

Kaufman - Trading Systems & Methods - Chap20

Chap20

VISIT...

LANZAROTE
Caliente.com

20 Advanced Techniques

Volatility is an essential ingredient in many calculations, from Wilder's RSI to variable stoploss points and pointandfigure box sizes. But volatility is more uniform than its shortterm measurement would imply; it is a more predictable component of a price series than the trend. Volatility is usually the main ingredient in risk; the more volatile the market, the greater the risk. As systematic programs mature, there seems to be a greater, justifiable concentration on how to manage volatility.

MEASURING VOLATILITY

In general, the volatility of most price series, whether stocks, financial markets, commodities, or the spread between two series, is directly proportional to the increase and decrease in the price level. This pricevolatility relationship has been described as lognormal in the stock market and is very similar to a percentagegrowth relationship. In Chapter 11 (PointandFigure Charting"), it was shown that soybeans increased in volatility at an average rate of 2.38% relative to price, very much the same as a logarithmic increase.

$$(n\text{-day volatility}) V(n) = c \times \ln (P_{\text{today}} - P_{\text{base}})$$

where $V(n)$ is the n -day volatility

P_{today} is today's price

P_{base} is the base price of the commodity, some

c is a scaling factor, near 1

This shows that, beginning at its base price, the volatility of a market increases in proportion to its price increase. To express this relationship for interest rate markets, it is necessary to use yield rather than price. In addition, as we have discussed from time to time throughout this book, that currency volatility cannot be expressed this way because it has no base price; instead, it has a point of equilibrium. Prices become more volatile as they move away from equilibrium in either direction.

Shifting the Volatility Base

It follows that prices are more volatile at higher levels and that most trading systems must cope with this change by adjusting their parameters. For example, a stoploss in gold might be \$2 per ounce when the price is under \$350 per ounce, \$4 per ounce at about \$1000, and \$10 per ounce at \$5000. A pointandfigure box size might vary in the same way as the stoploss.' as prices become higher, it requires a larger box size to maintain the same frequency of trend changes and signals. Similarly, a swing chart will need a larger reversal criterion.

Using the stoploss as an example, most traders are willing to take a fixed amount of risk (for example, \$500), regardless of whether this risk is too large or too small for market conditions; for many traders, it is only a

matter of how much they can afford to lose. A stoploss that is based on margin offers some improvement because margins are set according to market risk and contract size., however, the lag time needed for the exchange to change the margin is far too long to keep this relationship current. A percentage stop is a popular

472

solution for analysts who realize that volatility increases with price, but it falls far behind during major bull and bear markets. A reasonable representation of longterm or underlying risk is the adjusted, lognormal pricevolatility relationship. Although volatility may vary greatly at any price level, this relationship establishes a foundation for the normal level.

Base Price

The volatility relationship must include the price level at which volatility is essentially zero. Of course, we cannot find that level on a chart because no trading would have occurred. Although we do not know the price now, we call the level at which volatility is zero the base price. All interest rates, stock indices, and commodities have a base price. Certainly, if Treasury bills were to decline to a point at which the yield was near zero, most investors would choose to place their money elsewhere, causing activity (and volatility) to disappear. Similarly, when the price of corn fails to \$1.75 per bushel, below the cost of production, farmers are not inclined to sell; they will wait until prices rise. This waitandsee approach reduces volatility

Figure 201a shows that the base price can be found by detrending the data and formulating the volatility based on the detrended values. Although Figure 201b does not indicate a time period, volatility only makes sense when measured over some interval. Once detrended, a scatter diagram of price versus volatility should show a relationship similar to Figure 201b. Figure 201c is a reminder that the magnitude of the volatility and the pricevolatility relationship, is directly proportional to the time interval over which the volatility is measured.

Considering the pricevolatility relationship of gold since 1976 gives a typical example of how to find the base price. Figure 202a is a scatter diagram of monthly gold prices versus the monthly change in price, taken as positive numbers. A more sophisticated study would look at the price range over the month, rather than the net price change. The range will approach zero a little more slowly than the price change. Note that there is a cluster of dots in the price range from \$100 to \$175 per ounce and then at \$300 to \$450 per ounce. The lower values can be related to the pre1980 price levels, while the higher grouping shows low volatility during periods of higher prices since 1980. Again, the use of monthly price changes can yield a value of zero even if there was significant volatility during the month, simply because prices ended at the same place at they began. Using a price range will show a smoother pattern of price versus volatility

Detrending the price of gold using the Consumer Price Index, a technique applied to gold in Chapter 3, improves the uniformity of the results, seen in Figure 202b. Instead of two separate clusters of dots at lowvolatility levels, there is only one cluster in the range from \$110 to

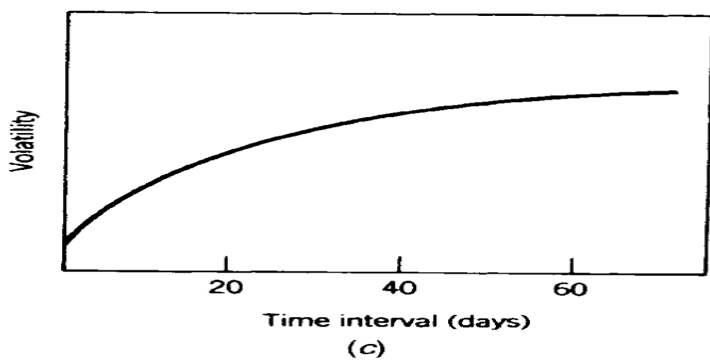
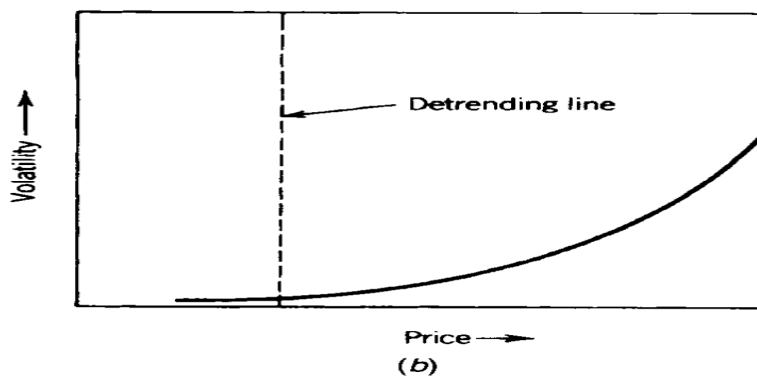
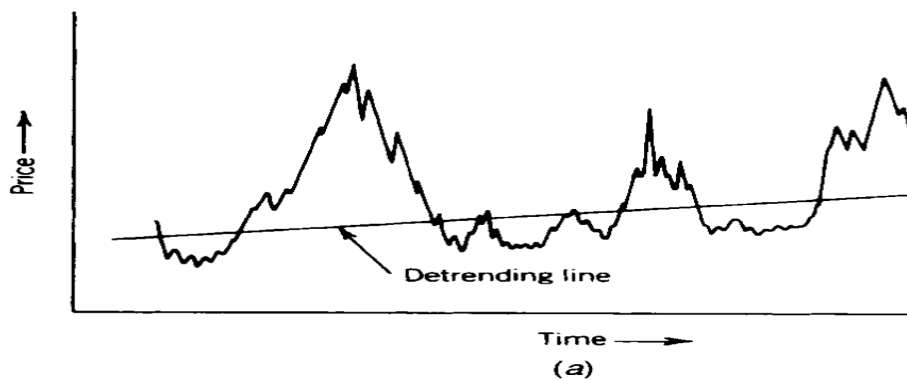
\$250 per ounce. A curved line has been drawn to represent the pattern of declining volatility in relationship to declining price. This line could be straight if prices were further adjusted using a log or exponential function. According to this curve, volatility approaches zero at a slower rate as prices drop below \$200 per ounce.

The time period over which volatility is measured is also a significant factor in the pricevolatility relationship. Longer measurement periods give higher volatility values. Regardless of the number of days used to determine volatility volatility will increase as price increases. This direct relationship will be very uniform for exchangetraded markets except when prices are very high. Exchange rules may limit trading when prices move too quickly; therefore, charts of real market prices will show that volatility stops expanding when it collides with these artificial constraints.

The magnitude of price movement will also increase as the period of measurement gets longer. During a volatile interval, prices will move farther in one direction during 3 or 4 days than they will in 1 or 2 days. As this interval gets very long, the volatility does not keep increasing at the same rate; it tends to slow, as shown in Figure 201c. The flattening

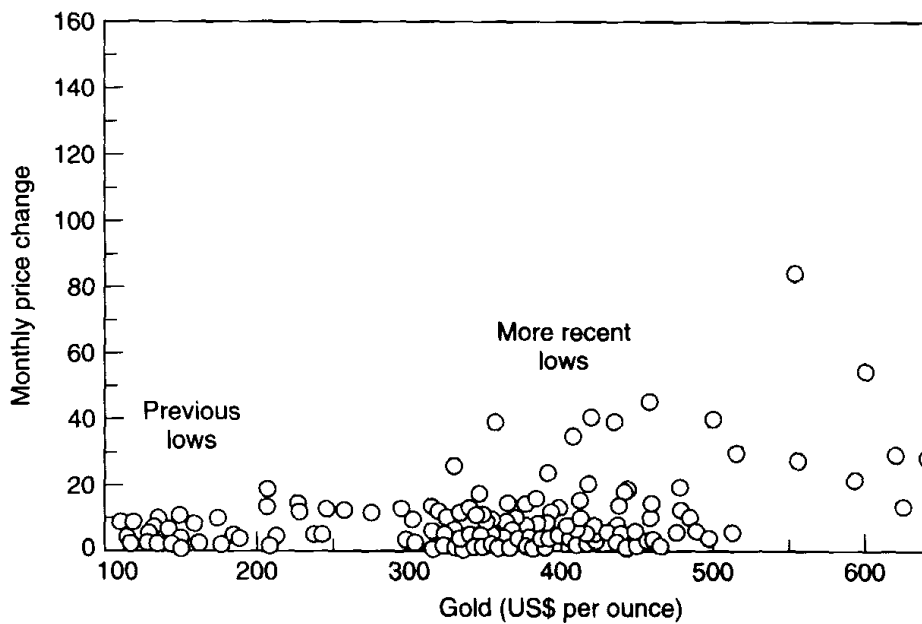
473

FIGURE 201 Measuring volatility from a relative base price. (a) Prices become less volatile relative to a longterm deflator (detrending line). (b) Volatility as a function of the detrended price. (c) Change in volatility relative to the interval over which it is measured.

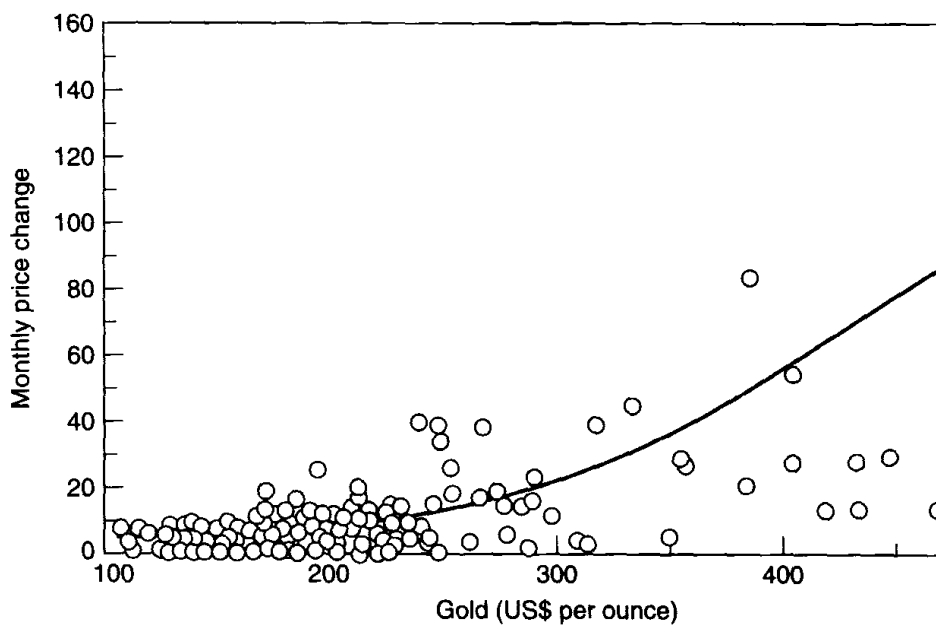


474

FIGURE 202 Finding the base price where volatility is zero, gold 1976-1993. (a) Volatility versus price. (b) Volatility versus deflated prices.



(a)



(b)

of the curve occurs at the point equal to the duration move.

Alternate Measures

The three most used measures of volatility all satisfy that is, they all expand at an increasing rate as price has been *net price change over time*. This is equivalent to P_{t-n+1}), where volatility is calculated over n days. The following is the probability of a price change at different levels

475

Markets that do not have a significant price change may also be volatile. Many systems use the daily or weekly highlow range to define risk and volatility. This technique can work, although it does not always capture a full measure of activity. The sum of the absolute price changes is a more descriptive value. Using Figure 203, the three measures can be shown as:

1. Price change over time interval n :

$$V_t = P_t - P_{t-n+1}$$

2. The maximum fluctuation during the interval n :

$$V_t = \max (P_t, P_{t-n+1}) - \min (P_t, P_{t-n+1})$$

3. The sum of the absolute price changes over the interval n :

$$V_t = \sum_{i=t-n+1}^t |P_i - P_{i-1}|$$

In (1), the volatility is entirely dependent on the value of ΔP regardless of the price activity that occurred during the time periods, a predictable change can be expected, but it is not as good as either (2) or (3).

The maximum range (2) corrects for the dependence on ΔP to produce a more consistent measure of volatility, which may be used as a method may be effective as the basis for a stop-loss, but it does not account for price fluctuation at each price level. Because prices end at the same level 20-2b, the close-to-close method (1) would show little volatility.

The sum of the absolute price changes (3) is the total volatility, although it is nondirectional. It clearly shows volatility in cases that are not apparent to either (1) or (2). A price that goes to highs alternately each day is much more volatile than a price that goes to the highs and then back to the lows only once during the day. This is useful for indicating a change in market character, and for

Ratio Measurements

Bookstaber¹ presents a volatility measurement V_t which is a ratio of prices, the high and low, or a combination of the two as follows,

where C_t is the closing price on day t

H_t is the high price on day t

L_t is the low on day t

V_t is the volatility on day t

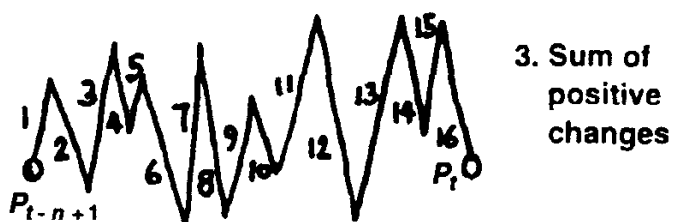
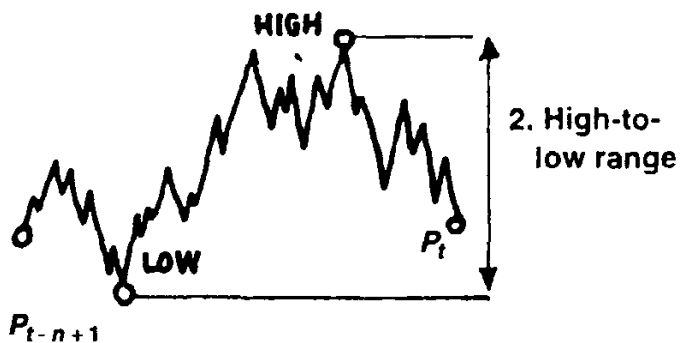
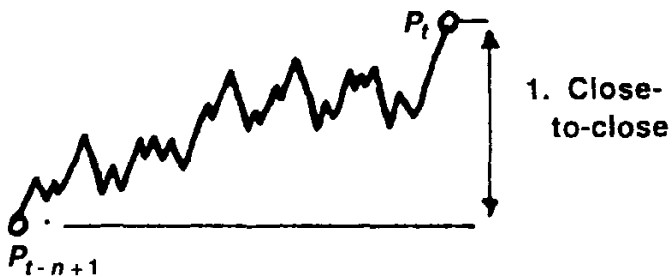
(a) *Close-to-close volatility*

$$R_t = \frac{C_t}{C_{t-1}}$$

and

$$A = \frac{1}{n} \sum_{i=1}^n \ln R_{t-i}$$

¹ Richard Bookstaber, *The Complete Investment Book* (Scott, Foresman, Glenview, IL, 1987).



then

$$V_t^2 = \frac{1}{n-1} \sum_{i=1}^n (\ln R_{t-1} - A)^2$$

and

$$V_t = \sqrt{V_t^2}$$

(b) *High-low volatility*

$$V_t = \frac{.601}{n} \sum_{i=1}^n \ln \left(\frac{C_t}{C_{t-1}} \right)^2$$

(c) *High-low-close volatility*

$$S_t^2 = .5 \ln \left(\frac{H_t}{L_t} \right)^2 - .39 \ln \left(\frac{C_t}{C_{t-1}} \right)^2$$

$$V_t^2 = \frac{1}{n} \sum_{i=1}^n S_{t-i}^2$$

$$V_t = \sqrt{V_t^2}$$

Note that the current t may be a time interval rather than a price of the period, and H_t and L_t are the highest and lowest

In the close-to-close estimation, the volatility is estimated using only the close-to-close price ratios. Bookstaber states that this measurement

the actual volatility during the current period t can be estimated by the distribution.

Volatility System

The daily volatility can be used in a trading strategy,

$$V_t(n) = \frac{1}{n} \sum_{i=1}^n D_{t-i+1}$$

where D_t is the maximum of:

- $|H_t - C_{t-1}|$
- $H_t - L_t$
- $|L_t - C_{t-1}|$

and H_t is the high on day t
 L_t is the low on day t
 C_t is the close on day t

Note that D_t is the *extended range* ("true range") of the day's work. V_t is the average extended range over the n days represented with this is

Buy if the close C_{t+1} rises by more than $k * V_t(n)$

Sell if the close C_{t+1} falls by more than $k * V_t(n)$

The volatility constant k is given as approximately 3, but it can be varied higher or lower to make the trading signals less or more frequent, respectively.

Volatility Filters

High volatility is clearly related to greater risk, but low volatility may also mean that there is no chance of profits. This is especially true for trendfollowing systems. The following are reasonable expectations for selecting trades based on volatility:

Entering on very high volatility will increase risk. It may be best to simply avoid those trades.

Entering on low volatility seems safe, but prices often have no direction and produce losses. Waiting for an increase in activity before entering might improve returns.

Exiting a position when prices become very volatile should reduce both profits and risk, but may come after the fact. This is an issue best resolved by testing.

Filter or Delay

Whenever a high or low volatility situation occurs at the time of an entry signal, there are two choices. The trade can be completely eliminated by filtering, or it can be delayed until the high volatility drops or the low volatility increases to an acceptable level. If the trade ends during the waiting period, the trade is eliminated. Both of these cases have been studied and the results follow.

Constructing a Volatility Fitter

Calculating the volatility is simple to program using any spreadsheet or strategy testing software. The following steps were used here:

Ibid., pp. 224-236.

478

- 1. Calculate a moving average trend.**
- 2. Calculate the volatility, using any one of the methods but not including the volatility of the current day.**
- 3. Enter a new trade if the volatility is (a) above the high-filter level.**
- 4. Exit a current position if the volatility is above the current price change is a profit or (b) the current price**

To give these choices a chance to show which are robust, five different markets were tested, each for more than 10 years ending in 1993: Eurodollars, Japanese yen, crude oil, IBM, and the S&P 500. Futures market prices were gap adjusted and indexed, using the nearest delivery month. Results from these tests are shown as percentage changes; IBM was quoted in share price. The trend speed (a simple moving average) was the same as the period over which the volatility was calculated. Periods of 35 days and 10 days were the only ones tested; 35 days was arbitrarily chosen as about one eighth of a trading year. The 10 day period (2 weeks) was included as a short term contrast.

Standard Deviation Measurement

A standard deviation was used to determine the volatility threshold level. For example, for a high volatility filter, a 1 standard deviation threshold means that no trades were taken if the volatility was above the average volatility plus 1 standard deviation, the top 16%. A 2 standard deviation threshold puts volatility in the top 2.5%, and a 3 standard deviation threshold means the top .5%. A 0 standard deviation would filter all trades above the average volatility. Because only 35 days are used, actual volatility can jump well beyond the normal 3 standard deviation

maximum. Filter values above 3 standard deviations (up to 7 standard deviations in these tests) must be used to isolate the most extreme volatile price movements.

Entry Filter Results

Table 201 shows the results for both high and lowvolatility entry options, expressed as rate of return based on maximum drawdown. The case where the standard deviation value is zero indicates no filtering; tests were run for a calculation period of 35 days. In part (a), the highvolatility filter causes a delay until volatility drops below the designated standard deviation level; in part (b), volatility must increase to be above the standard deviation level.

Results for the highvolatility filter are not impressive. The Eurodollar improves for lower levels while the Japanese yen and crude oil do not. Both IBM and the S&P improve by waiting for lower volatility, but both have underlying losing strategies; therefore, omitting trades is likely to improve results. Because these results are a function of both returns and risk, the elimination of high volatility does not clearly improve either component of performance.

The lowvolatility filter is much clearer. In all cases except the S&P, taking only those trades that occur with conditions of higher volatility shows improvements for all markets. Because of the few cases that would have occurred with volatility greater than 6.0 standard deviations, the large return shown by IBM is not realistic. A standard deviation between 1 and +1, indicating the practical elimination of 16% to 84% of all trades, shows the most consistency.

HighVolatility Exits: Distinguishing a Volatile Good Move from a Bad One

it may be that exiting due to high volatility is acting afterthefact; then, it would be ineffective for reducing risk. if a market is noisy then price jumps are shortlived, in which case you should close out positions when a volatile move is profitable because profits will soon disappear. Or, if the volatile move is an immediate loss, then waiting should recover at least

StDev	Euro\$	Yen	Crude	IBM	S&P
<i>(a) High-Volatility Entry Filter with Delayed Entry</i>					
none	835	340	218	13	-6
6.0	886	340	182	7	-4
5.0	876	340	183	9	-4
4.0	884	361	183	18	-4
3.0	932	399	176	19	-17
2.0	948	401	197	12	-17
1.0	988	410	198	20	-22
0	923	425	323	-35	-27
<i>(b) Low-Volatility Entry Filter with Trade Elimination</i>					
6.0	109	266	139	3,181	-40
4.0	457	47	94	308	-55
2.0	1,172	85	-20	816	-85
0	491	340	83	214	-95
-1.0	860	337	156	-19	-67
-2.0	882	320	219	19	-36
none	835	320	84	13	-36

*Thirty-five days' calculation period, rate of return based on maximum drawdown.

part of the loss. The opposite is true for a trending market. A good price move should be followed by more profits, and a bad move by further losses.

Tests of exit filters were separated into positive and negative price moves. When the positive option was elected, exits occurred when the highvolatility level was reached and prices generated a profit for the day. When the negative was used, prices must have produced a loss for the day. For the Eurodollar, returns improve when volatile losing moves are liquidated, and returns decline when profitable moves are closed out. We can conclude that the Eurodollar is a trending market, and cutting a trade short due to increased volatility is not a good strategy. That is not true with the S&P, which has a high degree of noise over short time periods. It was also not the case for the yen, which has erratic price behavior. Both of these markets improved when profits were taken on a positive move.

Because shorterterm analysis operates in an environment of greater noise, the exit filters will generally be an improvement. For markets with significant longterm trends and systems that are intended to capitalize on those trends, exiting a trade for almost any reason, other than a change of trend, is going to adversely affect the performance.

Creating Your Own Filters

A volatility filter is a simple calculation. Deciding which volatility filter to use is more complex because it requires a number of different testing programs. The following Omega program and the spreadsheet program (Table 202) combine all of these features into one testing program. An option is used to select each feature.

OMEGA Easy Language Code.

{High and Low Volatility Filters

Copyright 1994-1998, P.J. Kaufman. All rights reserved.

Do not enter trades on high volatility and price in trend direct
and price change of direction option.

480

TABLE 202 Spreadsheet Sample
Corel Quattro Spreadsheet Code

Col	Description	Begin Row	
L3	Low-filter limit factor		[Constant]
M3	High-filter limit factor		[Constant]
A	Date	5	[Input]
B-D	Open, High, Low prices	5	[Input]
E	Closing price	5	[Input]
F-G	Vol, Oplnt data	5	[Input]
H	Positive value of price changes	6	@ABS(E6..I
I	35-day moving average	39	@AVG(E5..I
J	Average volatility	39	@AVG(H5..I
K	Standard deviation of volatility	39	@STD(H5..I
L	Low-filter limit	39	+J39-\$L\$2*
M	High-filter limit	39	+J39+\$M\$2*
N	Position	40	@IF(I40<I39
O	Low-volatility indicator	40	@IF(H40<L
P	High-volatility indicator	40	@IF(H40>M
Q	Buy/sell with low filter	40	@IF(\$N40=
			@IF(\$H40>
			@IF(\$N40=
			@IF(\$H40>
R	Buy/sell with low delay	40	@IF(\$N40=
			@IF(\$N39=
			@IF(\$H40>
S	Buy/sell with high filter	40	@IF(N40=1
			@IF(\$H40<
			@IF(\$N40=
			@IF(\$H40<
T	Buy/sell with high delay	40	@IF(\$N40=
			@IF(\$N39=
			@IF(\$H40<
U	Exit on high volatility	40	@IF(\$H40>

Option 0 = No entry or exit filter

Option 1 = Entry filter only

Option 2 = Exit filter only - no price direction

Option 3 = Exit filter only - profitable move

Option 4 = Exit filter only - contrary move

Option 5 = Both entry and exit filters (options 1 and 2)

input: option(2), length(35), Efactor(1.0),ELfactor(1.0)

vars: vavg(0), vsd(0), lowlimit(0), Euplimit(0), Xuplimit(0),
totalpl(0),

risk(0), ratio(0), position(0), variance(0), change

{Volatility = average + standard deviation of price change

Calculate average and sd before new prices}

change = @AbsValue(close - close[1]);

vavg = @Average(change[1],length);

vsd = @StdDev(change[1],length);

{Extreme volatility limits}

lowlimit = vavg - vsd;

Euplimit = vavg + 2*Efactor*vsd;

```

Euplimit = mavg + 2*Elfactor*vstd;
Xuplimit = mavg + 2*Xfactor*vstd;

[System rules based on moving average trend]
mavg = @average(close,length);
if mavg>mavg[1] and position = -1 then position = 0;
if mavg<mavg[1] and position = 1 then position = 0;

[Enter on new trend signals below high volatility:
Efilter or new trend signals above low volatility: ELfilter]
if option=1 or option=5 then begin
[High volatility delay]
[ if mavg>mavg[1] and position <> 1 and change<Euplimit then begin]
[Low volatility filter]
[ if mavg>mavg[1] and mavg[1]<=mavg[2] and change<Euplimit and change>ELuplimit then
[Low volatility delay]
    if mavg>mavg[1] and position <> 1 and change<Euplimit and change>ELuplimit then b
        buy on close;
        position = 1;
        end;
[High volatility delay]
[ if mavg<mavg[1] and position <> -1 and change<Euplimit then begin]
[Low volatility filter]
[ if mavg<mavg[1] and mavg[1]>=mavg[2] and change<Euplimit and change>ELuplimit then
[Low volatility delay]
    if mavg<mavg[1] and position <> -1 and change<Euplimit and change>ELuplimit then b
        sell on close;
        position = -1;
        end;
    end;

[Normal entry only once in the same direction]
if option<>1 and option<>5 then begin
    if mavg>mavg[1] and position <> 1 then begin
        buy on close;
        position = 1;
        end;
    if mavg>mavg[1] and position <> -1 then begin
        sell on close;
        position = -1;
        end;
    end;

[Don't reenter after high volatility exit]
if option>1 then begin
    if mavg>mavg[1] and change<Xuplimit and position <>1 then begin
        buy on close;
        position = 1;
        end;
    if mavg<mavg[1] and change<Xuplimit and position <> -1 then begin
        sell on close;
        position = -1;
        end;
    end;

[Exit without volatility]
if option=0 or option=1 then begin
    if mavg<mavg[1] then exitlong on close;
    if mavg>mavg[1] then exitshort on close;
    end;
[Exit on high volatility and profitable direction]
if option=2 or option=3 then begin

```

```

    if mavg<mavg[1] or (close>close[1] and chang
    if mavg>mavg[1] or (close<close[1] and chang
end;
{Exit on high volatility and not a profitable d
    if option=2 or option=4 or option=5 then begin
        if mavg<mavg[1] or (close<=close[1] and chan
        if mavg>mavg[1] or (close>=close[1] and chan
    end;

{Replace high volatility days because they dist
{ if change > vavg+2*factor*vsd then change = v

```

Note that parts of the program have comment braces around one or more lines. These represent other options that may be tested. it will be necessary to remove the comment braces from one section and add them to another.

TRADE SELECTION

Once you have a basic system, the next step is to decide whether you can eliminate some of the losing trades without eliminating the profitable ones. In the extreme, we want to eliminate all of the losing trades, even if we reduce some of the profits of course. that is impossible. Every system has a risk, even socalled riskless trades. Arbitrage. when done properly, has virtually no risk; however, it may be so competitive that the opportunities are rare and the margin of profit small. You might find, after months of trading a guaranteed arbitrage, that your returns are disappointing compared with the cost of doing business. Putting your money into U.S. bonds would have been nearly as good with a lot less work.

Nothing is free. Good trading is not a matter of hitting it rich on a single short position: it is grinding out a profit day by day and week by week. Occasionally, if you follow new markets, you might find an opportunity that others have not yet seen. If you act fast, you can capitalize on this for a short time, until others see the same situation. Then bigger players come in and push you out. In turn, they are followed by other investors who combine to remove all the profit opportunity, even for themselves.

When selecting trades to eliminate, the easiest place to begin is by associating performance with volatility or price level. While some systems perform better in an environment of higher volatility, it may be that the best return relative to risk occurs when there is less volatility In general,

markets with extremely low volatility do not perform well because they have less direction. Under moderate volatility conditions, systems will return both lower profits and smaller losses; with high volatility we can expect large profits and high risk.

Predicting Volatility with Trading Ranges

William Brower³ performed a thorough study of trading ranges, range for tomorrow. This would be important for day traders and higher volatility is an advantage. Specifically, high volatility is desirable. A summary of his results is shown in Table 20-3 for the S&P 12/15/95.

Thomas Bierovic, OnBalance True Range
To visualize the change in volatility, Thomas Bierovic has created OnBalance True Range by following the same rules as OnBalance Volume, but using the True Range calculation instead of Volume. He then calculates a 9day exponential smoothing of the OnBalance True Range and uses the crossovers of the Oscillator and smoothed oscillator to confirm signals. Although the highs and lows may come at nearly the same time as other oscillators.

³ William Brower, Inside Edge (Inside Edge Systems, 10 Fresenius Road, Westport, CT 06880, March April 96),

TABLE 203 Predicting the Trading Range of the S&P 500*

Rule	Conditions Tested	
1	$O < L[1] - x$	Very good predictor, but
2	$O > H[1] + x$	Modest predictor, but fe
3	$C[1] + x > O > C[1]$	Range got smaller when
4	$C[1] - x < O < C[1]$	Range got smaller when
5	$O < C[1] - x$	Range tended to increas
6	$H[2] \leq H[1]$ and $L[2] \geq L[1]$	No significance
7	$H[2] > H[1]$ and $L[2] < L[1]$	Modest predictor of low
8	Day of week	Monday had lowest vola
9	$Average(TrueRange,3)[1] > x$	Higher avg true range fo good predictor of high
10	$RSI(close,3)[1] < x$	When RSI < 40 good pr

* Source: William Brower.

the positioning of the relative peaks and valleys may offer new patterns. For many traders, this simple interpretation can help separate high and low volatility conditions.

PRICE VOLUME DISTRIBUTION

Many systems that have excellent historic simulations fail when they are actually traded. These include models based on both intraday and daily data. One reason for this disappointing performance is the lack of understanding of market liquidity. Consider two systems:

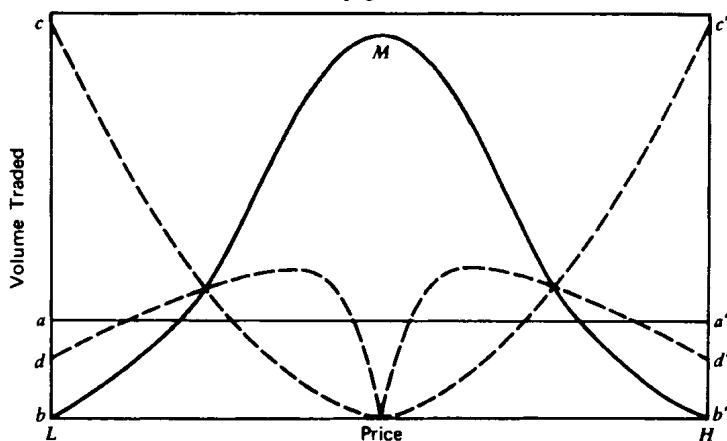
A trendfollowing method, which signals a buy or sell order when prices rise or fall relative to a specific price during the day

A countertrend system, which sells and buys at relative intraday highs or lows

Although each system intends to profit with their opposing philosophies, both act when prices make a relatively unusual move. Figure 204 shows the normal distribution of intraday volume and the profitability associated with this distribution.'

4 Mary Ginter and Joseph J. Ritchie, "Data Errors and Profit Distortions," in Perry J. Kauffman (Ed.). *Technical Analysis in Commodities* (John Wiley & Sons, New York, 1980).

FIGURE 204 Daily price/volume distribution.



484

The solid line W is the actual volume on a theoretically normal day in which prices remained in a trading range. Volume is greatest at the median M and declines sharply to the high and low endpoints H and L, where only one trade may have occurred. The dotted line cc' represents the apparent profit for a countertrend system that makes the assumption of a straightline volume distribution, ad. The endpoints are shown to contribute the largest part of the profits when, in reality, no executions may have been possible near those levels. Assuming the ability to execute at all points on the distribution bb', the approximate profit contribution is shown as dd'. Trades are most likely to be filled at points that limit immediate profits.

For trendfollowing systems, no profits should be expected when buy or sell orders are placed at the extremes of the day. The actual price distribution W is the maximum that could be expected from such a system; in reality, the first day is usually a loss.

TRENDS AND NOISE

Throughout this book there are techniques that try to find the

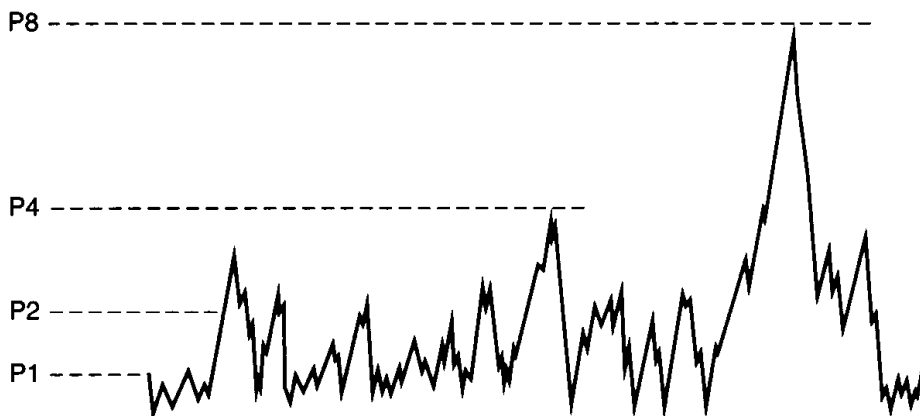
underlying trend when the price direction is not clear. The investor who buys sooner is more profitable, but only if the market moves higher. What makes it so difficult to identify the trend is noise. Noise is the erratic movement of price; by definition it is unpredictable, yet it has very definite statistical attributes.

Noise is the product of the market participants buying and selling at different times for different purposes. Each has their own objectives and time frame. Noise can also be a price shock an unexpected event that may or may not cause a lasting price change. Noise has most of the qualities of a sequence of random numbers. Nearly half of all price moves change direction from the previous move; about 25% of all price moves continue in the same direction for two periods, and so on. These patterns should already be familiar to analysts. The size of a price move, or price shock, also appears to be random; that is, 50% are quite small, 25% are twice as large, 12.5% are four times as large, and a few are very large.

Then how can you tell if a price move up is an indication of a new direction or simply more noise? The answer is that you can't tell by studying only price until after the direction has continued. There are situations, of course, where the news will give us information to show that the price jump was a structural change (for example, when the Federal Reserve lowers interest rates). But price alone won't tell us.

We can improve our chances of selecting the trend direction based on a net price move by choosing only the largest moves. This is the reason why a longterm trend is more reliable than a short one. To see this clearly, consider Figure 205 to be random numbers, or market noise. Most of the time prices remain below the level P1, but a price move that reaches P2 (twice as large) is still fairly common. Less often, price will jump to level P3 but only rarely will it reach P5. Let's say that, out of 1,000 price moves, P1 is reached 500 times, P2 occurs 250 times, P4 occurs 64 times and P5 about 4 times. If we use a breakout system to decide the price trend, and our breakout criteria is less than P2, then there would have been 250 total signals (both long and short) in which to find the right trend direction. If there were 10 confirmed trends during this period, you needed to reject or filter 240, or 96%, of all moves. Suppose your criteria was P5 instead of P2, and because the size of the subsequent price trend must be bigger due to the larger entry criteria, there was only one good trend during this period. Then you have a 1 in 4 chance of being correct, far better than the 1 in 25 using a very small trend criteria.

It is the magnitude of the market noise that determines the criteria needed for entering a trend. For a broadly traded market, the noise pattern is very predictable, while markets that have low volume might not conform closely to a traditional statistical profile.



Noise can be measured using the same calculations as volatility, given at the beginning of this chapter, or as Kaufman's efficiency ratio found in Chapter 17 ('Adaptive Techniques').

EXPERT SYSTEMS

An expert system is one in which you draw conclusions based on an accumulated knowledge base, which has been stored as data, facts, and relationships in the form of if-then-else rules. For our purposes, these would include price and economic data as well as the knowledge that, for example, highly correlated markets move together, and high volatility means high risk. When the first expert systems were developed, the core information was actually gathered by interviewing experts, which is how the name expert system was derived. The success of this method depends upon the quality and completeness of the knowledge base.

The knowledge base is used by an inference engine, which is able to draw conclusions from the facts stored in the knowledge base. Therefore, if we have the following information,

FACT 1: U.S. bonds are more volatile than Eurodollars.

FACT 2: The S&P 500 is more volatile than U.S. bonds.

FACT 3: Volatility is directly proportional to risk.

then the inference engine can create the new fact

FACT 4: The S&P is riskier than Eurodollars.

The inference engine provides a straightforward, logical process; however, there may be many relationships to resolve, not all of which may apply to the problem you would like to solve. For example, if we also have the following facts:

FACT 5: The S&P has a high degree of noise.

FACT 6: Eurodollars have a low degree of noise.

FACT 7: The S&P is currently trading below its 200-day moving average.

FACT 8: The S&P is currently trading above its 20-day moving average.

FACT 9: The S&P has been above its 200-day moving average for 80% of the past 20 years.

FACT 10: Eurodollar rates are driven by monetary bank policy.

FACT 11: Current monetary policy is dominated by concerns of inflation.

FACT 12: Inflation results in higher interest rates.

486

FACT 13: Net trading returns for the S&P have been 40% per annum.

FACT 14: Net trading returns for the Eurodollar have been 20% per annum.

then human logic might conclude that Eurodollar yields are likely to rise because of concerns over inflation. In addition, the S&P has greater risk, erratic behavior, and is currently not as strong as it has been on average over the past 20 years, although it is now rising. Compared with Eurodollars, there is greater risk trading the S&P but its returns have been higher.

How can this expert conclusion be duplicated by a computer? By adding a set of rules that parallels the thinking of experts, a computer can theoretically arrive at the same conclusions. For example,

RULE 1a: IF the Producer Price Index rises by more than the annualized rate of 4%, THEN we have inflation.

RULE 2a: IF there is high noise, THEN there is less chance of a trend.

RULE 3a: IF there is high noise, THEN there is greater risk.

RULE 4a: IF the faster trend is above the slower trend, THEN prices are trending up.

For each positive rule 1 through 4, there should also be a negative rule:

RULE 1b: IF the Producer Price Index does not rise by more than 4% annualized, THEN we do not have inflation.

RULE 2b IF there is low noise, THEN there is a greater chance of a trend.

RULE W IF there is low noise, THEN there is lower risk.

RULE 4b: IF the faster trend is below the slower trend, THEN prices are trending down.

Even with the negative rules there are some ambiguous cases. For example, in Rule 4 there is the case in which the two trends are in conflict. In other situations, the positive rule might be true, but the negative rule may not be as strong. In Rule 1b, we see that the effect is that "we do not have inflation" rather than we have deflation."

Forward Chaining

The process of combining the rules and facts to yield an expert opinion is called forward chaining. For example, beginning with FACT 6, "Eurodollars have a low degree of noise," we find the relevant rule, RULE 3b: IF there is low noise, THEN there is lower risk," and create a new fact: "Eurodollars have (relatively) low risk." Note that in each case, the terms low, high, and faster are all relative.

Once we have this new fact, that Eurodollars have relatively low risk, and we similarly conclude that the S&P has relatively high risk, we can also conclude that Eurodollars have lower risk than the S&P. The process of following the path of each fact as it is handled by various rules is called forward chaining. It will lead to other rules and other facts; it may be that the expert opinion will be found along this route, or that the combination of new facts, such as the relative risk between the Eurodollars and the S&P, will provide the expert opinion.

This example shows only a few facts that can be easily

summarized., however, there are thousands of pieces of information about performance characteristics, relationships to other markets, and fundamental factors that might alter expectations. If accumulated by asking experts, it is also likely that there will be conflicting information. While the human brain has a remarkable ability to sort through these items and select the information it considers most relevant, some important items can be overlooked when there is too much to

487

consider. An expert system is expected to use all of the data and reduce it to a single decision. In doing this, it must also select the most significant facts and resolve conflicts associated with the proper order of events and the time horizon of the investor.

A Technical Expert System

An expert system can treat indicator values as expert opinions and create systems without the fundamental relationships shown in the previous section. An example by Fishman, Barr, and Loick,' applied to the DJIA, defines rules as the relationships between the various indicators and calculations, and the facts as the values of those items. The purpose of this expert system is to inspect trending and nontrending characteristics of price movement to give the probability of a continued trend. The following example is adapted from their article:

Rule 1: IF $ADX > 18$ AND $ADX \geq ADX[21]$

(AND nontrending is false or undefined),

THEN there is a 95% chance the market is trending.

Rule 2: IF $ADX < 30$ AND $ADX \leq ADX[2]$

(AND trending is false or undefined),

THEN there is a 90% chance the market is nontrending.

Rule 3: IF the market is trending,

THEN $MACD = \text{probability of } .8$ AND $SD = \text{probability of } .5$ (85%).

Rule 4: IF the market is nontrending,

THEN $MACD = \text{probability of } .55$ AND $SD = \text{probability of } .75$ (75%).

Rule 5: IF stochastic AND $SK[11 < 30$ AND $SD[1] < 30$ AND $SK[1] < SD[11]$

AND $SK < SD$, THEN sell with 80% confidence.

Rule 6: IF stochastic AND $SK[1] > 70$ AND $SD[11 > 70$ AND $SK[11 > SD[11]$

AND $SK > SD$, THEN buy with 80% confidence.

Rule 7: IF $MACD$ AND $MACD \text{ value} > \text{signal line value}$, THEN buy with 75% confidence.

Rule 8: IF $MACD$ AND $MACD \text{ value} < \text{signal line value}$, THEN sell with 75% confidence.

where the following periods are assumed:

ADX is an 18period ADX calculation

$ADX[2]$ is the ADX 2 periods ago

$MACD$ is the Moving Average Convergence/Divergence with smoothing constants .15 and .075

SK is a 9period, %K fast stochastic

SD is a 9period, %D slow stochastic and undefined means that there is no information about the values. The facts are the actual values to be used in the evaluation:

FACT 1: ADX = 19

FACT 2: ADX[2] = 19

FACT 3: SK = 90

FACT 4: SD = 80

'Mark B. Fishman, Dean S. Barr; and Walter J. Loick, 'Artificial Intelligence And Market Analysis,' Technical Analysis of Stocks & Commodities (Bonus Issue 1993).

488

FACT 5: SK[11] SK=68

FACT 6: SD[I] SD = 92

FACT 7: signal line value = 70

FACT 8: MACD value = 68

FACT 9: the market is trending

The results, expressed as probabilities, offer greater insight into the likelihood of success using this method. Arriving at these probabilities, however, requires additional decisions. Without any further information, we can assume that there is only a 50% chance that the stochastic is a correct trend indicator. if we were to test the number of times the stochastic indicated an upward move with the number of days that the subsequent price moved higher, we could get a much better indication of the chance of success.

FUZZY LOGIC

in the second example of expert systems, the results were expressed as probabilities. Because trend and indicator calculations are estimations of market movement, very little is strictly true or false; therefore, probabilities are a more realistic way to view the results. The probability of being right can be represented by the reliability (percentage of profitable trades) of a system using these calculations. When right, the average payout is the average profits per trade, and when wrong, the loss is the average loss per trade. When you can assign a probability to results, the variable has often been calledfuzzy. But this way of looking at values is really a twist on basic probability analysis.

The idea of fuzziness is intended to describe the lack of precision in normal human conversation and thought. In its pure form, the use of fuzziness allows human uncertainty to be introduced into methods of artificial intelligence. For example, we often say,
"Therewere a lot of people in line at the show."

"I had to wait a long time."

"It was really cold while I was waiting."

"The stock market was strong yesterday."

"Unemployment dropped sharply."

in all these cases, we understand what is being said although there are no specific values associated with "a lot of people," "a long time," "cold," "strong," and "sharply" In fuzzy logic, all is not true or false, 0 or 1, there or not there. True fuzzy logic will answer the question "if a halfeaten

apple is still an apple, how much do you have to eat before it stops being an apple?"

The concept of fuzziness includes fuzzy numbers, such as "small," "about 8," "close to 5," and "much larger than 10," as well as fuzzy quantifiers, such as "almost," "several," and "most." Phrases such as "Unexpected results of Government reports cause big moves" is a common fuzzy expression.

Fuzzy Reasoning

Fuzzy events and fuzzy statistics are combined into fuzzy reasoning. It is a remarkable phenomenon that the answers to the following examples are clear to the human brain, but not to a machine (the answers can be found at the bottom of this page).'

6 Parts of this section are drawn from Perry Kaufman, Smarter Trading (McGrawHW, NY, 1995).

.pvq (~) p is (Z) '11,fa (T) :am saldumxa axp ol siamsui! ;)U _

489

Example 1: X is a small price move. Y is much smaller than X. How small is Y?

Example 2: Most price moves are small. Most small price moves are up. How many price moves are up?

Example 3: It is not quite true that the quarterly earnings were very bad.

It is not true that the quarterly earnings were good.

How bad were the quarterly earnings?

Common Approach to Fuzzy Solutions

In the terminology of fuzzy logic, there are three aspects of problem solving: membership, to show how the data are related to each other; fuzzy rules, to draw conclusions; and defuzzifiers to turn the fuzzy answers back into useable results.'

For example, we would like to predict whether a price move will be big enough to capture a profit needed by our trading system. To have a robust system, we always use a 20day trend but would like to predict which price moves will be above our minimum needs before we enter the trade. In general, we have found that, given commissions and slippage, we need to net \$500 per trade' Because we are using a trendfollowing system that gives up profits before exiting, we would like to see an \$800 profit to capture a net of \$500. Based on our needs, we can define our limits as:

1. Average peak move between \$300 and \$800 is acceptable
2. Average peak move over \$800 is desirable.
3. Average peak move under \$300 is unacceptable.

To decide whether current price movement is likely to give us enough profit to enter a trade, we measure the average price move over every 20day period for the past 1 year (longterm) and for the past 1 month (shortterm). Based on this approach, we create the rules:

1. if both the longterm and shortterm profit potentials are unacceptable, then the current profit potential is unacceptable.
2. If both the longterm and shortterm profit potentials are desirable, then the current profit potential is desirable.

3. If both the longterm and shortterm profit potentials are acceptable, then the current profit potential is acceptable.

These three cases are very clear, but what if the longterm and shortterm expectations are different? For example, if the longterm average peak profits are \$250 and the shortterm

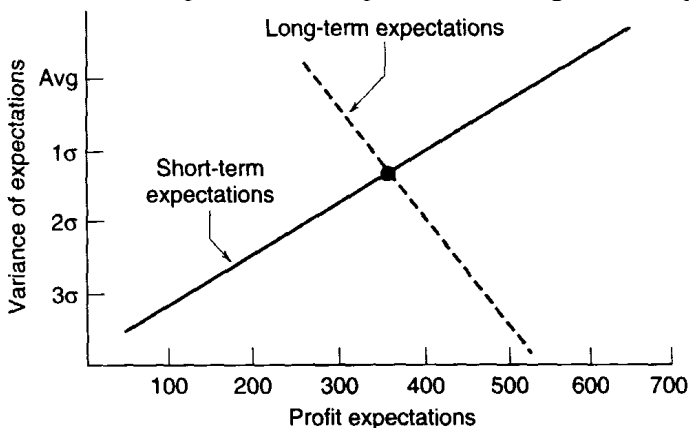
are \$600, what can be expected from the current move? In the current application of fuzzy logic, this converts to a problem in probability for which we need more information about the history of these price moves. Through testing, we find that the 1 year average price move has a standard deviation of \$75 and the past 1 month has a standard deviation of \$200. We could then construct a diagram that shows the expected results from combining the two measurements (see Figure 206). The average returns are shown with descending lines based on their standard deviations. These cross at about \$350 per trade, giving a reasonable answer to how to combine the two values.

One additional measurement that cannot be overlooked is the potential error in each statistic. The longerterm measure used 250 trading days, while the shortterm used only

'Murray A. Ruggiero, Jr., "Artificial trader jumps candlesticks," *Futures* (February 1995), and "Lighting candlesticks with fuzzy logic," *Futures* (March 1996).

490

FIGURE 206 Expectations of a price move using two time periods.



about 20 days. The standard error for each calculation is 6.3% and 22.3%, respectively. In the final interpretation of results, we can say that the longerterm data are more important than the shortterm values by about 3.5 :

1. Then we could write additional rules:
4. if the longterm potential is unacceptable and the shortterm is acceptable, then the current potential is unacceptable.
5. If the longterm potential is unacceptable and the shortterm is desirable, then the current potential is acceptable.

Using just the rules, we can claim that we are applying Fuzzy logic to our trading; however, the state of the art indicates that fuzzy logic is still hanging onto the coattails of simple probability. When it lets go, it is likely to be the most significant breakthrough in market ~is.

FRACTALS AND CHAOS

Another new area that has captured the interest of market analysts is chaos. Chaos theory is a way to describe the behavior of nonlinear systems, those that cannot be described by a straight line. Chaotic systems are not necessarily without any form or method, as the expression is commonly used, but those phenomena that are not a simple variation of a linear relationship. Some examples of this behavior will be given later in this section.

one method of measuring chaotic systems is with various geometric shapes. This effort has resulted in an area of mathematics now called fractal geometry; its approach strikes a true note about how the real world of numbers actually works.' All of us have been taught Euclidean geometry in school; it is the world of straight lines and clean edges in which we can measure the length of a line or the area of a rectangle very easily. in the real world, however, there are no straight lines; if you look closely enough, they all have ragged edges and all may be described as chaotic.

Fractal Dimension

in fractal geometry we find that there is a way of representing the irregularity of numbers, and the formations seen in nature. We first must accept the notion that there are no whole numbers in nature, that realworld objects are more likely to be described as fractional, or having a fractal dimension. The classic example of this is the algebra of coastline dimen

9 An excellent discussion of this topic can be found in Edgar Peters, Chaos and order in the Capital M~ (John Wiley & Sons, New York, 1991). The coastline example is originally credited to Benoit Mandelbrot,

491

sion. We will see that the questions "How long is the coastline?" and "How far did prices move?" are very similar.

The answer to both these questions is "That depends on how it is measured. Consider the problem of measuring the coastline of Australia using a large wall map. if we take a 12inch ruler, we might find that the coastline is about 10 feet (perhaps 10,000 miles according to the scale). Had we taken a slightly smaller ruler we would have been more accurate and perhaps have found the coastline to be 11,000 miles; even a smaller ruler would have followed the contours better and found 12,000 miles of coast. As the ruler gets smaller the coastline appears to get longer. If we had an infinitely smaller ruler, the coast would be infinitely long. There is really no correct answer to the question, "How long is the coastline?"

Fractal dimension is the degree of roughness or irregularity of a structure or system. In many chaotic systems, there is a constant fractal dimension; that is, the interval used for measuring will have a predictable impact on the resulting values in a manner similar to a normal distribution. Therefore, using a '121nch ruler, we measure the coastline and get 10,000 miles:

24inch ruler	5,000mile coastline
12inch ruler	10,000mile coastline
6inch ruler	20,000mile coastline

Note that the large 24inch ruler returns a value that is actually smaller than what we believe is a reasonable answer. This is because, when you place a long ruler from one point to another on the map, it cuts across part of the land mass.

Using Fractal Efficiency

In Chapter 17 ('Adaptive Techniques'), there is a discussion of Kaufman's efficiency ratio. This ratio is formed by dividing the net change in price movement over n periods by the sum of all component moves, taken as positive numbers, over the same n periods. If the ratio approaches the value 1.0, then the movement is smooth (not chaotic); if the ratio approaches 0, then there is great inefficiency or chaos. This same measurement has more recently been called fractal efficiency.

Fractal efficiency measures the amount of chaotic movement in prices; this can also be considered similar to market noise. In Chapter 17, Kauffman related this to trending and nontrending patterns, when the ratio approached 1.0 and 0, respectively. While each market has its unique underlying level of noise, the measurement of fractal efficiency is consistent over all markets. Markets may vary in volatility, although their chaotic behavior is technically the same; therefore, the characteristics of one market may be compared with others by matching both the fractal efficiency and volatility.

Interpreting fractal efficiency as noise allows other trading rules to develop. For example, a market with less noise should be entered quickly using a trending system, while it may be best to wait for a better price if the market is classified as having high noise. A noisy market is one that continues to change direction, while an efficient market is smooth. These characteristics are important in choosing trading rules and in turning a theoretical model into a profitable trading system.

Chaotic Patterns and Market Behavior

Chaotic patterns are easy to imagine in the behavior of prices, but very difficult to measure. There would be no problem in predicting price direction if every participant reacted in the same way to the same event, much the way a single planet would smoothly orbit a single sun. In the real world nothing is quite as simple. Consider the pattern of prices represented as planets that are affected equally by two events, E_1 and E_2 . We will get a wobbly pattern whenever the planet passes across the midpoint where one attractor is stronger

492

than the other, shown in Figure 207a. At point a , the object is most affected by the nearest attractor, E_1 , but as it circles it becomes closer to E_2 and tries to form an orbit around it. The possible patterns are too complex and they vary based on the distance between E_1 and E_2 and the size of E_2 compared with E_1 . If attractor E_2 is much larger than E_1 , there will simply be a distortion in the orbit around E_2 ; if E_1 and E_2 are the same size, objects a and b will switch orbits, forming figure eights.

As complex as these patterns might get, they are simple when compared with reality. Each day brings events of various importance into the market, acting as attractors. Each attractor has an initial importance that

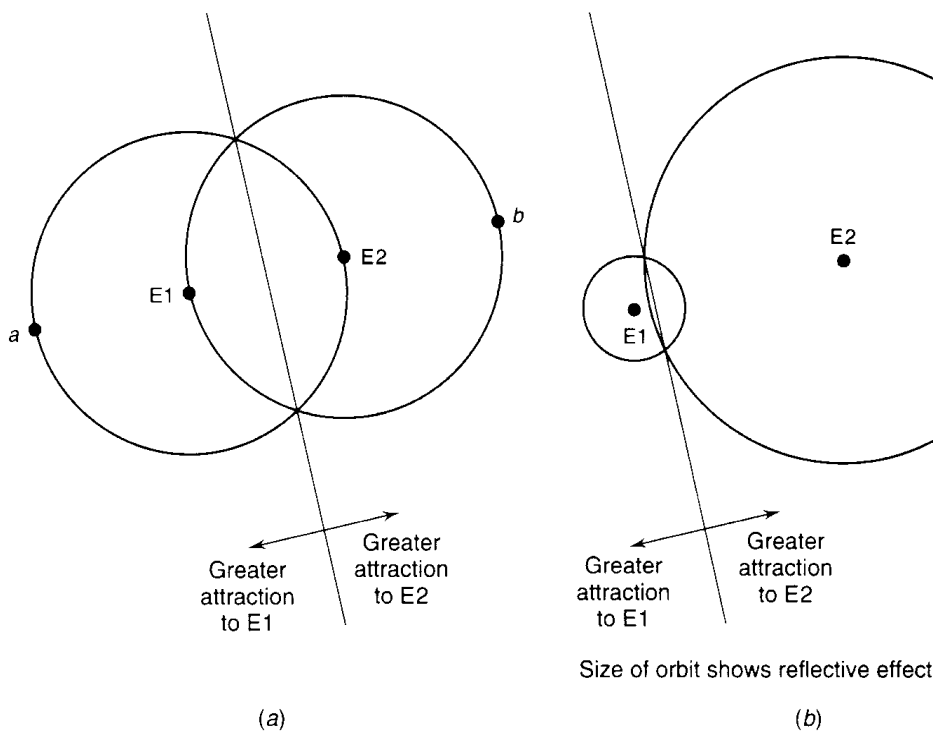
loses value over time. To make matters worse, we cannot predict when a new attractor, or news event, will appear. This makes the chaotic pattern very similar to raindrops falling on a pond. Each new drop, equivalent to a new event, hits at an unpredictable time and place, with variable size, and forms circular ripples. These ripples dissipate as they get further away from the point of contact in the same way that the importance of an event fades away over time. The interesting aspect of the raindrop analogy is that, while we cannot predict where the next raindrop will fall, once it has landed we can completely determine its effects until the next drop hits. This is remarkably similar to the market. Less often there is a significant event, a price shock, that overwhelms the smaller events, the market noise, for a short time.

NEURAL NETWORKS

Another area of analysis that has continued to grow rapidly is neural networks. During the past 5 years there has been a deluge of material on the advantages of using neural net

Parts of this section are based on Perry Kaufman, *Smarter Trading* McGrawHill 1995, pp 16411)~

FIGURE 207 (a) Equal attractors cause a symmetric pattern that switches between E 1 and E2. (b) Very unequal attractors show only a small disturbance in a regular orbit.



works as a tool for uncovering market relationships and finding better systems. Most of what has been said is true. This technique offers

unparalleled power for discovering nonlinear relationships between any combination of fundamental information, technical indicators, and price data. its disadvantage is that it is potentially so powerful that, without proper control, it will find relationships that exist only by chance.`

Although the idea and words for the computerized neural network are based on the biological functions of the human brain, an artificial neural network is not a model of a brain, nor does it learn in the human sense. It is simply very good at finding patterns, whether they are continuous or appear at different times.

The operation of an artificial neural network can be thought of as a feedback process, similar to Pavlov's approach to training a dog:

1. A bell rings.
2. The dog runs to one of three bowls.
3. If right, the dog gets a treat; if wrong, the dog gets a shock.
4. If trained, stop; if not trained, continue at step 1.

Terminology of Neural Networks

The terminology used in the computerized neural network is drawn from the human biological counterpart, shown in Figure 208. The principal elements are:

Neurons, the cells that compose the brain, process and store information.

Networks, groups of neurons.

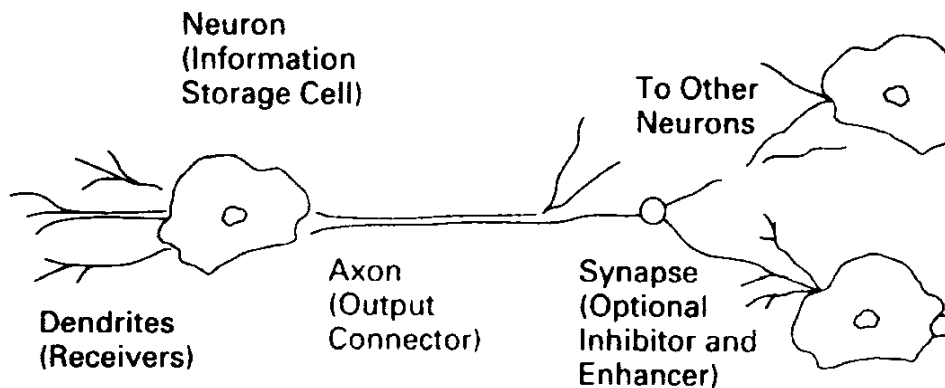
Dendrites, receivers of information, passing it directly to the neurons.

Axons, which come out of the neuron and allow information to pass from one neuron to another.

Synapses, which exist on the path between neurons and may inhibit or enhance the flow of information between neurons. They can be considered selectors.

Readers are referred to E. Michael Azoff, *Neural Network Time Series Forecasting of Financial Markets* (John Wiley & Sons, 1994), for a thorough treatment of this subject, and Edward Gately, *Neural Networks for Financial Forecasting* (John Wiley & Sons, 1996), for a more introductory approach.

FIGURE 208 A biological neural network. Information is received through dendrites and passed to a neuron for storage. Data are shared by other cells by moving through the output connector, called an axon. A synapse may be located on the path between some individual neurons or neural networks; it selects the relevant data by inhibiting or enhancing the flow.



Source: Perry Kaufman, *Smarter Trading* (McGrawHill, 1995, p. 165).

494

Artificial Neural Networks

Using essentially the same structure as a biological neural network, the computerized, or artificial neural network (ANN), can generate a decision on the direction of the stock market. It relies heavily on the synapses, which are interpreted as weighting factors in this process. To achieve its result, it will also combine inputs that interact with one another into a single, more complex piece of data using layers of neurons, as shown in Figure 209, a classic 3layer neural network.

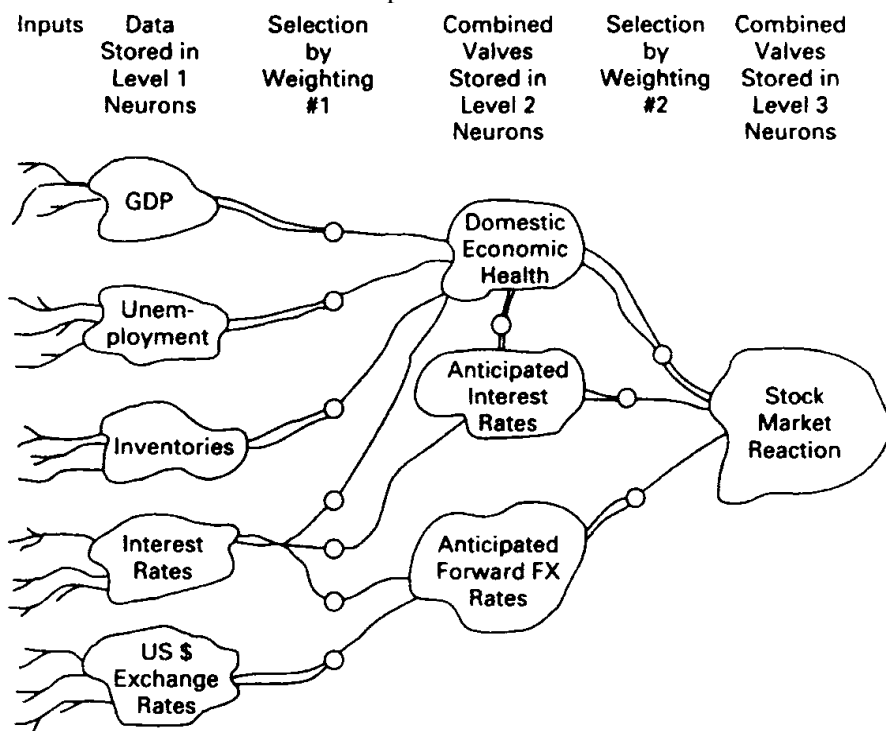
To be most efficient, the inputs to the ANN should be those factors considered most relevant to the direction of stock prices. The five items chosen here were the Gross Domestic Product, unemployment, inventories, interest rates, and the value of the U.S. dollar—all readily available data. Each of these items is input and stored in separate neurons. Changing values may have a positive or negative effect on the final output, which is the direction of stocks. An improved GDP, lower unemployment, higher inventories, and a lower U.S. dollar are all good signs for the economy and result in the possibility of higher interest rates (defensive action by the Federal Reserve to prevent inflation) and a likely decline in stock prices. Interest rates themselves have a direct effect on stock prices, improving profits as rates decline.

Each neuron in layer 1, which receives the data from the dendrites, is connected to a second layer of neurons through a synapse. Each synapse can be used to filter or enhance information by assigning a weighting factor. For example, if changes in unemployment have a greater impact than changes in the GDP, then it may receive a weight of 1.5 compared with .9 for the GDP. If very small changes in any of the data is considered unimportant to the result, then the synapse can act as a threshold and only allows data to pass if it exceeds a minimum value.

The second layer of neurons is used to combine initial data into significant subgroups. In Figure 209, the GDP, unemployment, and inventories are combined into a single item.

FIGURE 209 A 3layer artificial neural network to determine the direction

of stock prices.



Source: Perry Kaufman, *Smarter Trading* (McGrawHill, 1995, p. 166).

495

called Domestic Economic Health. The synapses allow each element to be assigned a specific level of importance using a weighting factor. Also note that this neuron is altered by anticipated interest rates, which is the result of data flowing to another neuron in layer 2. Finally, the three neurons in layer 2 are combined according to importance, giving the net stock market reaction to the input data.

The human brain works in a way very similar to the artificial neural network shown in Figure 209. It groups and weighs the data, combining them into subgroups and finally producing a decision. The human process of weighting the data is complex and not necessarily transparent; that is, we may never know the precise flow of data, how many layers exist, and how weights are assigned and reassigned before the final decision.

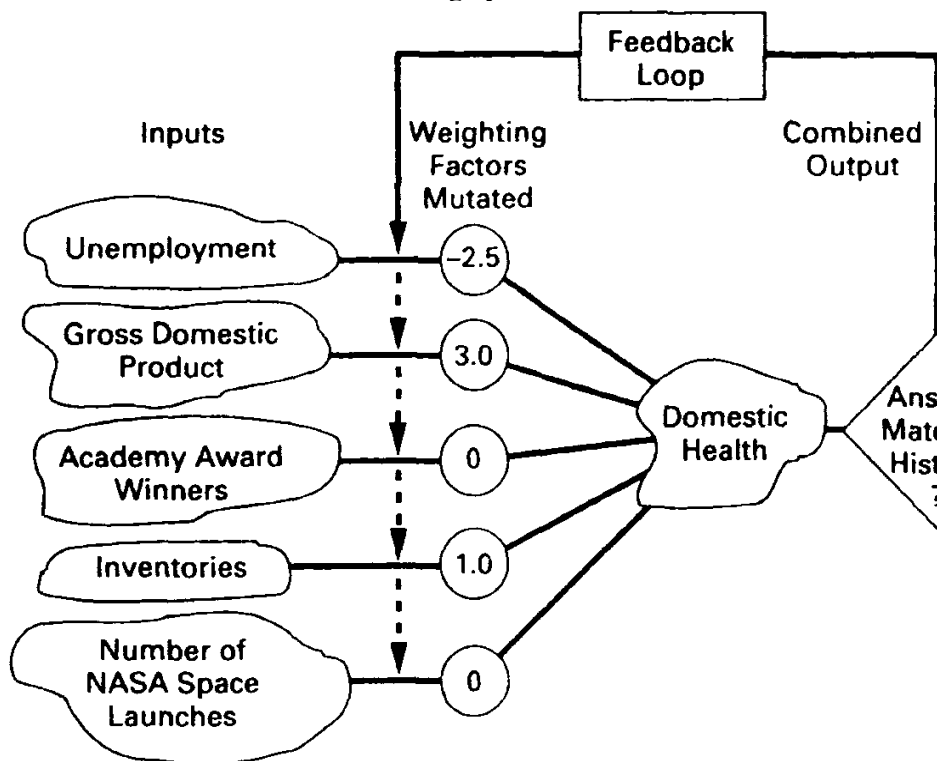
A computerized neural network is not as complex. Because it cannot know whether its answer is correct, you must tell it. This is done by giving the computer the historic data and the corresponding answers. By giving the artificial neural network a long history of information, it can determine, using feedback, the weighting factors that would have given the correct results most often. The more history that is given to the computer, the more likely it will find a robust answer. Figure 2010 shows the results of using five different inputs to predict the direction of the stock market.

Two of the inputs, the Academy Award Winners and the Number of NASA Launches are not likely to be useful in the long term, but may appear to provide valuable information for short intervals. By using enough comparisons, the weighting factors are found to show that unemployment has a strong negative effect on prices, the GDP a strong positive effect, and inventories have a weak positive effect. The other items had no consistent predictive ability and received a weight of zero. This feedback process is called training.

Selecting and Preprocessing the Inputs

There is considerable debate over the inputs needed to find a successful solution using a neural network; however, everyone agrees that the selection of inputs is critical. These inputs must be presented in the most direct form because, unlike an expert system, the

FIGURE 2010 Learning by feedback.



Source: Perry Kaufman, Smarter Trading (McGrawHill, 1995, p. 166).

neural network will not be able to change them. This step is called preprocessing. We must decide what factors affect the direction of stocks and the ability to anticipate that direction, then prepare data that contains information with these qualities. For example, we may want to know the shortterm and longterm trends., the direction of interest rates and the Dow Utility Index; the ratio of interest rates to gold; economic data such as the

GDP and balance of trade; technical indicators such as the RSI, stochastic, and ADX; and a 20day moving correlation between the U.S. stock market and other major markets. There are countless factors that might influence the direction of stocks; the more you choose, the slower the solution and the greater the chance of a less robust model. If you choose too few, they may not contain all of the information necessary; therefore, the preprocessing problem requires practice. You may also construct a number of systems that produce profits and include their basic components as inputs to the neural network. You might create a performance series for a specific system that has only values 1, 0, and -1, representing short, neutral, and long market positions. In that way the neural net may be used to enhance an existing trading strategy.

Selecting the Output

In a manner similar to evaluating optimization results, it is necessary to select the success criteria. This should be based on a combination of frequency of trading, the size of the profits per trade, a reward/risk criterion, and a frequency of profitability. According to Ruggiero,¹ neural networks can be used to predict price direction, such as the percentage change 5 days into the future; however, they are much better at predicting forward-biased technical indicators, because these tend to have smoother results.

Most technical indicators, such as trend and momentum calculations, smooth the input data, which is most often the price. The longer the time period used in the calculation, the greater the smoothing. Ruggiero suggests an output function such as

```
@Average((close[-5] - @Lowest(close[-5],  
@Lowest(close [-5],5)),5)
```

which is similar to a smoothed 5period stochastic shifted forward by 5 periods. In this notation, the bracketed value [5] represents 5 periods into the future.

The Training Process

At the heart of the neural network approach is the feedback process used for training, shown in Figure 20.10. To observe sound statistical practice, it is advised that only about 70% of the data be used for training. As the neural net refines the weighting factors for each of the inputs and combinations of inputs (in the layers between the input and output), it will need more data to test these results. Of the remaining 30% of the data, 20% should be designated for testing. The success or failure of the method to find a solution is based on the performance of the test data. The remaining 10% is saved for out-of-sample validation.

Weighting factors are found using a method called a genetic algorithm, discussed in the next section. For now, we need to know that the training process begins with an arbitrary or random value assigned to the weighting factors of each input. As the training proceeds, these weighting factors are randomly mutated, or changed, until the best combination is found. The genetic algorithm changes and combines weighting factors in a

manner referred to as "survival of the fittest," giving preference to the best and discarding the worst.

² Murray A. Ruggiero, Jr., "Build a real neural net," *Futures* (June 1996), the first of a series of excellent articles on this subject.

497

Testing is completed when the results of the test data converge to a single value. That is, after a number of feedback loops, the test data is used with the new weighting factors. If the results are improving the process continues. If the results are improving at a very slow rate, or have become stable at one level, the neural network process is considered completed. Sometimes the results get worse, rather than better. This can be fixed by beginning again using new random weighting factors. If this doesn't work, then the inputs must be reevaluated for relevance.

Because the genetic algorithm is a trialanderror process, rather than an analytic approach, the best results could be found by coincidence, rather than by cause and effect. With enough data series it is always possible that two series of events will appear related, although they are not. It is necessary to review the results to avoid simple mistakes

A Training Example

We would like to train a neural net to tell us whether we should buy or sell stocks. As inputs, we select what we believe to be the five most relevant fundamental factors: GNP unemployment, inventories, U.S. dollar index, and shortterm interest rates. This test does not use any preprocessed data, such as trends or indicators. To simplify the process, the following approach is taken:

1. Each input is normalized so that it has values between +100 and 100, indicating strength to weakness, with 0 as neutral.
2. When the combined values of the five indicators exceed +125, we will enter a long position; when the combined value is below 125, we will enter a short.
3. Combined values between +125 and 125 are considered neutral.

To show how the training process works, two events are shown in Tables 204 and Table 205 as Case 1 and Case 2. Table 204 is the initial state of the neural network, where we have chosen to set all of the weighting factors to 1.0. In actual training, the network might require that the sum of all weights total to 1.0. The actual values of the normalized inputs are shown in the columns marked "relative value," and the correct historic answer is at the bottom, marked as "strong" and "weak" stock market reactions to these values. For the neural network to return the correct answers, it must produce a result greater than +125 for Case 1 and below 125 for Case 2. By assigning initial weights of 1.0 to all inputs, the value of Case 1 is +75 and the value of Case 2 is +401 both fail and testing continues searching for better weighting factors.

TABLE204 Two Training Cases (Initial State)

Case 1					
Input	Relative Value		Weight	Net	Relative Value
GNP	Strong	50	1.0	50	Weak
Unemployment	Low	-25	1.0	-25	High
Inventories	Low	-50	1.0	-50	Neutral
U.S. dollar	Vry Strg	75	1.0	75	Neutral
Interest rates	Falling	-25	1.0	-25	Rising
Total value		75			
Threshold		+/-125			+/-125
Current response		None			None
Actual market reaction		Strong			Weak

Source: Perry Kaufman, *Smarter Trading* (McGraw-Hill, 1995)

498

TABLE205 Two Training Cases (After Mutated Weighting Factors)

Case 1					
Input	Relative Value		Weight	Net	Relative Value
GNP	Strong	50	1.0	50	Weak
Unemployment	Low	-25	-1.0	25	High
Inventories	Low	-50	1.0	-50	Neutral
U.S. dollar	Vry Strg	75	1.0	75	Neutral
Interest rates	Falling	-25	-1.0	25	Rising
Total value		125			-125
Threshold		+/-125			+/-125
Current response		None			None
Actual market reaction		Strong			Weak

Source: Perry Kaufman, *Smarter Trading* (McGraw-Hill, 1995).

In Table 205, the weighting factors have undergone mutation using a genetic algorithm. This example attempts to use only weighting factors of +1.0, 0, and -1.0. By reversing the effect of unemployment and interest rates on the direction of the stock market, and by random selection of weighting factors, the results now match the historic pattern of stock movement. Using fractional values for weighting factors, and many more training cases, the ANN method should find the current underlying relationship between these inputs and stock market movement.

Reducing the Number of Decision Levels and Neurons

The robustness of a neural network solution is directly related to the number of decision levels, the number of neurons in each level, and the total number of inputs. Fewer elements produce more generalized and, therefore, more robust solutions. When there are many decision layers and

many neurons, then the inputs can be combined and recombined in many different ways allowing very specific patterns to be found. The more specific, the greater the chance that the final solution will be overfitted, or that the neural network program will fail because it cannot converge on a satisfactory result.

The tradeoffs using a neural network are the same as most other optimization methods. Too many inputs and combinations require lengthy testing and a greater chance of a solution that is overfit. Too few values can produce a result that is too general and not practical. It is best to begin with the most general and proceed in clear steps toward a more specific solution. In this way you will understand the process better, ultimately save time, and be able to stop when you have reached the practical limitations of this method.

GENETIC ALGORITHMS

The concept of a genetic algorithm is based on Darwin's theory of survival of the fittest. In the real world, a mutation with traits that improve its ability to survive will continue to procreate. This has been applied to the process of finding the best system parameters. Although a genetic algorithm is actually a sophisticated search method that replaces the standard optimization, it uses a technique that parallels the survival of the fittest. It is particularly valuable when the number of variables is so large that a test of all combinations is impractical. Instead of the typical sequential search, it is a process of random seeding.

¹³ This technique is attributed to John H. Holland *Adaptations in Natural Language and Artificial Systems* (University of Michigan Press, 1975).

499

selection, and combination to find the best set of trading rules. Standard statistical criteria are used in the selection process to qualify the results.

Representation of a Genetic Algorithm

Using the words common to this methodology, the most basic component of a genetic algorithm is a gene; a number of genes will comprise an individual, and a combination of individuals (and therefore genes) is a chromosome. Each chromosome represents a potential solution, a set of trading rules in which the genes are the specific values and calculation expressions, and these form individuals that represent rules and ultimately form a trading strategy. For example, Chromosome 1 might be a rule to buy on strength:

If a 10day moving average is less than yesterday's close and a 5day stochastic is greater than 50, then buy

Chromosome 2 could be a rule that buys on weakness:

If a 20day exponential is less than yesterday's low and a 10day RSI is less than 50, then buy

If we rewrite these two chromosomes in a notational form, the genes and individuals in its structure become more apparent:

Chromosome 1: MA, 10, <, C, [0], &, Stoch, 5, >, 50, 1

Chromosome 2: Exp, 20, <, L, [1], &, RSI, 10, <, 50, 1

Each of these chromosomes has 11 genes, any one of which can be

changed. In addition, each chromosome has two individuals, separated by an "&" operator. In Table 206, the description of the genes indicates other values that can replace the current ones. Table 206 is a way to represent the chromosomes and individuals in a general form. It is easy to see that each gene can be changed, and each change will represent a new trading rule. A combination of trading rules, or chromosomes, will create a trading strategy. Before continuing, the following steps will be needed to use the genetic algorithm to find the best results:

1. A clear way of representing the chromosomes, or individuals
2. A criterion to decide that one chromosome is better than another
3. A selection procedure that determines which chromosomes will survive, and in what manner
4. A process for mutation (introducing new characteristics) and crossover (combining genes) to procreate chromosomes with greater potential

TABLE 206 Functional Description of the Genes in Chromosomes A and 2

Gene	Chr 1	Chr 2	Function
1	MA	Exp	A trend calculation (moving average, I)
2	10	20	The calculation period for the trend c
3	<	<	Relational operator (<, <=, >)
4	C	L	Price used in trend calculation ((H + L
5	0	1	Reference data ([0] = current day, [1]
6	&	&	Method of combining individuals (and
7	Stoch	RSI	Indicator calculation
8	5	10	Indicator calculation period
9	>	<	Relational operator
10	50	50	Comparison value for relational oper
11	1	1	Market action (1 = Buy, -1 = Sell)

500

Fitness

Having represented the chromosome in Table 206, we next must define a fitness criterion, which ranks the results. Because fitness will lead to survival, it is very important to decide which chromosomes should be discarded and which should be used further. A fitness criterion must combine the most important features associated with a successful trading strategy:

1. Net profits, or profits per trade
2. The number of trades, or an error criteria
3. The smoothness of the results

Ideally, we prefer systems that have large profits, lots of trades, and very consistent performance. To measure that result, the following might be used:

Rank = Profit per trade * (1 / sqrt(number of trades)) * (gross profits / gross losses)

If the profits per trade are larger, the rank is also higher; if there are more trades, the rank is higher, and if ratio of gross profits to losses is

greater, the rank is higher. These factors are not weighted, but they serve as a simple criterion for comparing or measuring the fitness of a chromosome. For example, a trading method that returned \$500 per trade for 10 trades, with a gross profit/loss ratio of .5 would have a rank of 170.75. Another system that returned only \$250 per trade over 100 trades with a profit/loss ratio of 1.4 would rank 315.00. Therefore, the system with the smaller returns has a much more agreeable trading profile according to the fitness criteria. Each analyst must create a criterion that allows those strategies with a personally desirable profile to survive.

Mutation and Crossover

Remembering that a genetic algorithm is a sophisticated search method, it needs a way to introduce new rules (genes), combine genes into individuals that have passed the fitness criteria, and discard genes and individuals that are less promising. This is done through mutation and crossover. Mutation is the process of introducing new genes, or combinations of genes, from a gene pool, which we have defined as a set of rules and relational operators, similar to those in Table 206. For example, the first gene, which represents a trendfollowing calculation, could be a moving average, exponential smoothing, linear regression, or breakout. In mutating the gene, one of these four techniques is chosen randomly. Similarly, the calculation period and the way in which rules are combined are selected randomly. Typically, only one of the individuals in the chromosome is mutated in each step; therefore, in Chromosome 1,

Chromosome 1: MA, 10, <, C, [0], &, Stoch, 5, >, 50, 1

the genes in the second individual (Stoch, 5, >, 50, 1) might be mutated to

Crossover is a way of combining two chromosomes

The result is called an *offspring*. Using chromosomes 1 and 2, two individuals to get

Offspring 1: Exp, 20, <, L, [1], &, Stoch, 5, >, 50, 1

Offspring 2: MA, 10, <, C, [0], &, RSI, 10, <, 50, 1

The two methods of mutation and crossover provide the only tools needed to introduce new features to the optimization and to combine features in all ways. If you could continue this process indefinitely, you can study every possible mutation and combination.

Selection

The process of natural selection is to allow only the best individuals to survive. A strong selection criterion chooses only those individuals with the highest ranking, as determined

by the fitness test. A weak criterion allows lower rankings to survive. The implementation of the selection process is also drawn from evolution.

When an individual has a high fitness score, it becomes a larger part of the population; when it has a low score, it is removed or reduced from the population. This can be done by modifying the list of conditions, which is used to create new genes and individuals randomly. For example, in the first gene we have the possibility of four trend criteria:

1. Moving average
2. Exponential smoothing
3. Linear regression
4. Breakout

Using random selection, each one has an equal chance of being picked. Let us say that we perform the first 10 tests without removing any of the possibilities, but only measure the average ranking of the results. On the 11 th test, we select linear regression and compare the results of its fitness with the average fitness of the first 10 tests. If the fitness is greater, we make a number of copies of this linear regression gene and add it to the list of trending methods. One popular way to calculate the number of copies is to divide the current fitness, by the average fitness, F , of previous tests. Therefore,

Number of copies = current fitness/average fitness

This method works best when there are a large number of choices rather than only four possibilities. However, if the fitness of the linear regression was 4, and the average fitness of previous tests was 2, we would make two copies, replacing the original entry. The new list would then have

1. Moving average
2. Exponential smoothing
3. Linear regression
4. Breakout
5. Linear regression

When a trend criterion is mutated, that is, another criterion is selected randomly from the list, there is now a 40% chance that a linear regression will be selected, and a 20% chance that another method will be used. Natural selection has resulted in the linear regression being more desirable because its fitness ranks higher. In the initial tests that formed the average fitness, there were other conditions that were mutated other than the trend criteria. It could be that these conditions were the reason for the higher score, and not the trend method. At the same time the first gene is copied, other genes that were part of the same chromosome were also copied, so that all of these genes have a better chance of appearing. Because we have not eliminated any of the other trend criteria, each one should appear from time to time. If the breakout method is actually better than the linear regression, it should have a higher fitness when it is mutated; it will then be copied and, theoretically, displace the linear regression as the survivor. Eliminating a choice may speed up the genetic algorithm, but introduces the possibility of removing a good technique on the basis that it only appeared in combination with other genes that were not strong.

Putting It into Practice: Training, Tuning, and Testing

The technique for a successful search using a genetic algorithm requires that the data be divided into three subsets: training, tuning, and testing in a manner similar to a neural network and using a procedure of step forward testing. Beginning with the first 1,000 data points, the system is trained; it is then tuned on the next 500 data points, and finally the performance using a test set of only 10 points is recorded. The process is then continued by shifting the data forward by 10 points. In this way, the performance is recorded for dis

502

joint tests of 10 points. This process implies that new parameters must be found every 10 data points for trading to duplicate the training method.

CONSIDERING GENETIC ALGORITHMS, NEURAL NETWORKS, AND FEEDBACK

Stumbling onto the perfect solution, or even a better solution, is always acceptable. Recognizing a good solution, by hard work or by chance, still requires talent many discoveries that seem to have occurred by chance are really the product of hard work that provides the opportunities for discovery. A genetic algorithm is a way of recognizing a situation that may be an improvement or a brilliant new method but not always. Unlike real life genetic mutations in which you can relate the change to the environmental need, this genetic algorithm may simply apply an isolated rule because of a single situation that may never occur again in the same way. To follow the path of a mutation, we need to confirm that this new rule is statistically and logically sound that there are enough cases to make it appear to be the right choice. Don't accept the results without careful thought and validation.

The large number of possible tests that make the genetic algorithm valuable also requires that you understand the concept of local versus global solutions. When you start with a random set of rules and parameters, then vary some and combine others, it is possible that you will find a local maxima, that is, a good solution but not the best. It is possible that a combination of trends and indicators that would have produced the perfect system was never seeded by the random process, nor was it found by changing some of the values during the process. One solution is to run the genetic algorithm a number of times to see if it arrives at the same solution starting with different random values. While it is not a guarantee that you will cover all major combinations, especially when there are a massive number of variations, it is the simplest way to avoid serious oversights.

Feedback is the basis for successful solutions using advanced searching methods, but it also creates uncertainty. For example, a neural network may be trained on 70% of the data and tested on 20%. When the test data is used to evaluate the weighting factors found using the training data, it provides feedback, which biases the selection of weighting factors in the training data. Although these data sets are separated, the constant feedback between the two provides a connecting tunnel that requires them to be considered as one data set. The testing data is no longer out of sample; therefore, another 10% of the data is withheld for scientific out of sample testing when the final solution is found. If that 10% data does not produce

results consistent with the training and testing, then the inputs must be changed and the process begins again. There is now a dilemma as to whether there is any outofsample data at all because the 10% held aside has been used as feedback for the entire process.

This problem is not unique to neural networks and genetic algorithms, but to all testing processes beginning with basic serial optimization. How do you test your results on current data without being guilty of feedback? There is really no answer. Statisticians claim that the best solutions use the most data; you can't overfit a solution when you test a large number of situations. The results are forced into being generalized. Holding a small amount of data aside for outofsample testing may not prove anything unless the results are exceptionally bad. If the outofsample results are poor, they may still be representative of one or two small periods during a long historic test. The final decision rests with the trader and not the computer. if the results seem reasonable, there was sufficient test data, the inputs were relevant, and the performance was monitored before trading begins, then the solution is likely to be good.